

# Open Research Online

---

The Open University's repository of research publications and other research outputs

## Discourse-centric learning analytics: mapping the terrain

### Journal Item

#### How to cite:

Knight, Simon and Littleton, Karen (2015). Discourse-centric learning analytics: mapping the terrain. Journal of Learning Analytics, 2(1) pp. 185–209.

For guidance on citations see [FAQs](#).

© 2015 The Authors



<https://creativecommons.org/licenses/by-nc-nd/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/4043>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

This is the final accepted draft (pre-copy editing) of:

Knight, S., & Littleton, K. (2015). Discourse-Centric Learning Analytics: Mapping the Terrain.

*Journal of Learning Analytics*.

Please cite the journal version of the paper.

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

## Discourse-Centric Learning Analytics: Mapping the Terrain

**Simon Knight**, Knowledge Media Institute, Open University, UK; [sigknight@gmail.com](mailto:sigknight@gmail.com)

**Karen Littleton**, Centre for Research in Education and Educational Technology, Open University, UK

There is an increasing interest in developing learning analytic techniques for the analysis, and support of, high quality learning discourse. This paper maps the terrain of discourse-centric learning analytics (DCLA), outlining the distinctive contribution of DCLA and outlining a definition for the field moving forwards. It is our claim that DCLA provide the opportunity to explore the ways in which: discourse of various forms both resources and evidences learning; the ways in which small and large groups, and individuals make and share meaning together through their language use; and the particular types of language – from discipline specific, to argumentative and socio-emotional – associated with positive learning outcomes. DCLA is thus not merely a computational aid to help detect or evidence ‘good’ and ‘bad’ performance (the focus of many kinds of analytic), but a tool to help investigate questions of interest to researchers, practitioners, and ultimately learners. The paper ends with three core issues for DCLA researchers – the challenge of context in relation to DCLA; the various systems required for DCLA to be effective; and the means through which DCLA might be delivered for maximum impact at the micro (e.g. learner), meso (e.g. school), and macro (e.g. governmental) levels.

**Keywords:** discourse, discourse-centric learning analytics, social learning analytics, data mining, computer supported collaborative learning, collaborative learning, learning analytics

### 1. INTRODUCTION

“Learning Analytics is an emerging research field and design discipline that occupies the ‘middle space’ between the learning sciences/educational research and the use of computational techniques to capture and analyze data (Suthers & Verbert, 2013).” (Knight, Buckingham Shum, & Littleton, 2014, p. 1). One interest for learning analytics is in its potential for the analysis of learning processes, and discourse data (Buckingham Shum & Ferguson, 2012). In 2014 we saw the fourth Learning Analytics and Knowledge (LAK) conference and the second Discourse-centric Learning Analytics (DCLA) workshop (Buckingham Shum, de Laat, De Liddo, Ferguson, & Whitelock, 2014) as a part of that conference. It now seems a good time to reflect on this developing field. In this paper, we discuss this sub-area of learning analytics Discourse-centric Learning analytics – which we refer to specifically in order to foreground our interest in the middle space between learning and analytics (Suthers, Lund, Rosé, Teplov, & Law, 2013; Suthers & Verbert, 2013); an interest not only in computational-analytic techniques for discourse (the ‘DC’ of DCLA) but in the explicit learning implications of those techniques, which should be grounded in our understanding of educationally salient discourse. This paper maps the terrain of the work falling under DCLA, describing the relationship between theory and methods, the various targets of analysis, and the kinds of learning claims at which those analyses might be targeted. We do not intend to put forward any particular stance with regard to the relationship between discourse and learning, but rather

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

by mapping the terrain we highlight necessary considerations in any well-motivated approach to DCLA as a distinctive research-area.

If learning analytics hopes to build on and contribute to learning theory, researchers should consider the ways in which learning analytics – in this case discourse-centric learning analytics – relates to, and is different from, other research on related topics. What is it that makes something DCLA, rather than natural-language processing, text based machine learning, or some other form of learning analytics? The interest in developing DCLA techniques is in part driven by recognition of its potential bi-directional contribution; as we discuss further below, DCLA might take two forms: bottom up (inductive, exploratory), or top down (deductive, confirmatory). In either case, we must consider how the general approach relates to a distinctive sub-discipline of DCLA, and its relationship to existing theory. While in this paper we are specifically interested in DCLA, we anticipate that similar issues arise for other sub-domains of learning analytics. We consider these challenges in the context of DCLA, concluding with three key questions relating to *context*, *systems*, and *feedback mechanisms*. The paper thus provides a mapping of the DCLA terrain and its distinctive contribution, providing a focal point for DCLA's key challenges. A key component of our claim is that DCLA provide the opportunity to explore: the ways in which discourse of various forms supports and evidences learning; the ways in which small and large groups, and individuals make and share meaning together through their language use; and the particular types of language – from discipline specific, to argumentative and socio-emotional – associated with positive learning outcomes. DCLA is thus not merely a computational aid to help detect or evidence 'good' and 'bad' performance (the focus of many kinds of analytic), but a tool to help investigate questions regarding the nature of discourse as learning, which are of interest to researchers, practitioners, and ultimately learners.

## 1.1 Preliminary Definitions for DCLA

Given its potential to support and investigate discourse in learning contexts, DCLA is thus of considerable interest. Whether in the context of Massive Open Online Courses (MOOCs), traditional university tuition, or more informal writing on the web, the automated or semi-automated analysis of linguistic data is increasingly prominent (see for example the classified sources in Table 1). Yet, it might well be noted that such analysis is not particularly new. Analysis of discourse and its communicative or illocutionary force is well established, and the ways in which new technologies mediate or alter such properties at least as old as Socrates' complaint regarding the diminished value of *written* argument (Plato, 1999). Given this context, it is important to consider: the ways in which *DCLA* might differ in focus from related fields; what its potential is; what perspectives are brought to bear on the consideration of DCLA as opposed to other domains; and what the defining features of DCLA are.

The trivial response of course is just that DCLA is the application of learning sciences, and analytic approaches, to student discourse data. However, insofar as this leaves unspecified definitional questions regarding 'discourse' 'learning analytics' and indeed 'centric', this is clearly inadequate. To give an example, imagine a case in which a learner's motivation is assessed during the writing of a formal

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

assessment (via non-linguistic proxies), and this motivation is fed back to them in some meaningful way. While certainly we might say this is learning analytics, and indeed we could accept that there is ‘discourse’ data involved (the written assessment), it seems less clear to us that we would wish to describe such analytics as Discourse-Centric Learning Analytics. This paper describes our developing perspective on this issue, and marks out the ground within which we think DCLA lies. In doing so, we describe a set of illustrative examples drawn from our own work at the Open University. These are not intended to be exhaustive, nor indeed exemplary, but simply exemplifications of the types of space within which we conceive DCLA, and the pressure points for such work.

A key focus of some earlier work, (see for example, Scheuer, Loll, Pinkwart, and McLaren’s 2010 review of the state of the art in computer-supported argumentation) has been the potential of automated analysis for discourse data – a key pivot point for DCLA. That review, and an earlier review (Clark, Sampson, Weinberger, & Erkens, 2007) of argumentation in Computer Supported Collaborative Learning (CSCL) environments, largely focus on existing argumentation frameworks – that is, they take the ‘top down’ theory driven approach. However, it is important to note the potential of DCLA approaches for contributing back to theory, in – well conceptualised – bottom up data driven approaches, as we discuss in the section ‘Top-down and Bottom-up Approaches’ below.

The first DCLA workshop (Buckingham Shum et al., 2013) proposed a mission statement for DCLA, to *“Devise and validate analytics that look beyond surface measures in order to quantify linguistic proxies for deeper learning.”* Yet, we hold the view that in a very real sense, linguistic activity is not simply an *indicator* (or proxy) for deeper learning, it is often the site of that learning. That is, our learning is not just *displayed* through discourse, the discourse forms a fundamental site of that learning. Indeed, **Error! Reference source not found.** below gives an indication of some focal points for DCLA, which apply across the varieties of data which might properly be considered ‘discourse’ (text and spoken-chat data, longer written monologues, CSCL interactions and so on). In each case, one can imagine – as is indicated in the columns – clear examples of educational foci, and benefits to such analysis. We highlight though, that as one moves down the rows, there is a shift in focus from individuals, to two types of small-group process (Stahl, 2010); in the first two – subject knowledge and rhetorical capabilities and tendencies – the emphasis is on individual capabilities and knowledge; in the third, it is on the ways in which people co-operate together, and share knowledge, this is the collective level; while in the fourth discourse is taken to be not only indicative of learning, but constitutive of that learning in a collaborate, co-constructive sense.

The purpose of the table is to highlight the multiple ways in which discourse may be probed, the different properties of such analysis, and the ways in which such analysis might interact. This is particularly important if, through our data collection, task setting, and analytic methods, we reify a particular perspective on the purposes of discourse which neglects the multi-functionality of discourse. This is true in both the context of summative and formative assessments of discourse quality; we should be careful to avoid giving feedback on only one aspect of discourse, of obscuring the nature of the discourse in attempts to reduce data to manageable, neat categories (enabling, for example,

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

visualization and dashboards), and of failing to appropriately scaffold both teachers and students in their use of talk for learning. In the latter case, alongside those concerns, the issue of performativity (see, for example, Ball, 1999) is raised – wherein those aspects of a curriculum which are directly assessed, become those aspects which are most focussed on in classroom contexts. Indeed (as with any assessment instrument) there is the risk that as certain forms of DCLA become widespread, teaching and assessment become driven by those facets of discourse that have been made visible by these techniques; that we value what we can measure, rather than measuring what we value (Wells & Claxton, 2002). Moreover, while quantification offers analytic opportunity, it is not the end goal of learning analytics properly understood; closing the loop for meaningful feedback is crucial to differentiating learning analytics from other approaches. This was the focus of the second workshop (Buckingham Shum et al., 2014), asking “*Once researchers have developed and validated discourse-centric analytics, how can these be successfully deployed at scale to support learning?*”. Understanding this full cycle is crucial, and it is to this issue that we turn in the section ‘The Learning Analytics Lifecycle’. While DCLA of various sorts may be productive, we should be aware of the interpretive flexibility of the tools we develop, and the pedagogic, assessment, and epistemological context they form a part of (Knight et al., 2014). However, if we are able to operationalize some stable and general categories and patterns in discourse use, then we can model such patterns in classifiers, recognising that such modelling involves a balance around acceptable levels of information loss (Rosé & Tovaes, In Press).

In the following section (‘The Learning Analytics Lifecycle’) we discuss how this process might occur, moving on to discuss the potential for ‘Top-down and Bottom-up Approaches’ in this analytics lifecycle, before then (in ‘2. DCLA – UNDERSTANDING WHERE DISCOURSE MEETS ANALYTICS’) discussing some specific examples of ‘2.1 Discourse Sites’ and ‘Analytic levels’, around which our concluding definition, and questions, are focussed.

**Table 1 – Discourse Functions, Focus, Content and example DCLA techniques\***

Function (from Knight, 2013)	Focus	Content	Example DCLA techniques
<i>Supporting individuals’ subject learning</i>	Advancement of subject knowledge	Propositional, semantic content of what is said. Appropriate reference to entities and relationships from the curriculum.	Techniques for recognising named entities, esp. ontology-driven techniques for domain specific entity-relationship extraction (e.g. history; biology) (Rosé et al., 2008; Tablan, Roberts, Cunningham, & Bontcheva, 2013)
<i>Supporting psychological development –</i>	Development of argumentation	The rhetorical content of language. How language is used to position and make	Techniques for extracting rhetorical and argumentation forms (Clark



Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

<i>the development of social and reasoning skills</i>	skills or dispositions	claims/discourse moves in discourse or writing	et al., 2007; De Liddo, Sándor, & Buckingham Shum, 2012; Scheuer et al., 2010; Weinberger & Fischer, 2006).
<i>Promoting group understanding or commonality</i>	The interactional nature of collaborative discourse, networks of documents	Who says what to whom? The actors constituting the interpersonal context	Techniques for identifying interlocutors in dialogue, and more diffuse social networks of strong/weak ties via different channels (Haythornthwaite & de Laat, 2010; Oshima, Oshima, & Matsuzawa, 2012; Sie et al., 2012)
<i>Enabling sharing of ideas that can be improved together</i>	Resources that emerge through the discourse, as ‘improvable objects’ (Wells, 1999).	The ways discourse – particularly in CSCL contexts – resources future discourse and artefact development. The temporal and contextual (including social) development of ideas.	The temporal and interactional nature of the above 3 rows, and the nature of the shifts between states – the <i>pragmatic</i> context. (Bannert, Reimann, & Sonnenberg, 2013; Furberg & Ludvigsen, 2008; Lid & Suthers, 2003; Medina & Suthers, 2009; Rosen, Miagkikh, & Suthers, 2011; Suthers & Desiato, 2012; Suthers & Medina, 2010).

\*Table extended from Knight and Littleton (2013).

## 1.2 The Learning Analytics Lifecycle

By an ‘analytics lifecycle’ we mean the cycle through which: a need or desire for an analytic technique is identified, the analytic itself is developed, and then implemented in some form. Consider the developed learning analytics lifecycle (Clow, 2012) which moves through: learners, data from or about learners, processing of data into metrics, and intervention. The process of ‘evidence centered design’ (Mislevy, Behrens, Dicerbo, & Levy, 2012) follows a similar pattern, beginning (1) with the identification of target constructs, then (2) identifying behaviours indicative of those constructs before (3) developing tasks likely to elicit those salient indicators. It is apparent, then, that DCLA should consider the types of tasks in which learners – formally or informally – engage, and the types of data that may be captured about those contexts. Theorising is needed to consider how to process this data into appropriate metrics – how to interpret the data. Once metrics are developed, consideration of interventions should focus on ways in which data might be meaningfully represented either for students or for expert educators in

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons 6  
 License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

support of learners (in line with Clow’s point that interventions do not necessarily need to involve a return of data to students).

Consideration of other broad models of analytic cycles including that of learning analytics (Elias, 2011) – as summarised in **Table 2** – indicate a broad alignment, in particular with the need to define goals, (learning goals in our context), in making decisions regarding data selection and capture. Each of the example cycles noted in the columns involves a set of key steps, from data selection through to use. It is important to note that even at the stages of selecting and processing data a range of irreducible elements involving theoretical models, target constructs, and data constraints are involved. For example, the segmentation of discourse data into chunks for processing is both a pragmatic decision regarding the selection of statistical technique, and also a theoretical one around the units at which data may be meaningfully discussed. Such claims position DCLA: Theory, in this case largely psychological in nature, gives insight into learning that can lend itself to development of analytic techniques, both with respect to a manual analysis and in tandem through provision of proof of concept tools for learning technologists. A crucial question to ask at each DCLA stage – from task design, data collection, and the steps of analysis, through to feedback to students – is “what is it for?”.

**Table 2 - Comparison of Analytics frameworks and models, (Elias, 2011, p. 10)**

Knowledge Continuum	Five Steps of Analytics	Web Analytics Objectives	Collective Applications Model	Processes of Learning Analytics
Data	Capture	Define Goals	Select	Select
Information	Report	Measure	Capture	Capture
Knowledge	Predict		Aggregate	Aggregate & Report
			Process	Predict
Wisdom	Act	Use	Display	Use
	Refine			Refine
		Share		Share

Given, then, the emphasis of various analytic models on selecting data, and defining goals, an interesting question is raised – to what extent can ‘bottom up’ approaches play a role in developing our theory and empirical understanding of phenomena, given the need for theory in data selection?

### 1.3 Top-down and Bottom-up Approaches

As we note in the introduction, the interest in DCLA is in part driven by recognition of its potential bi-directional contribution; DCLA might take two forms: bottom up (inductive, exploratory), or top down (deductive, confirmatory). Either might appropriately address a learning analytic development cycle – feeding into how we understand our desired construct, the types of behaviours associated with it, or the kinds of feedback likely to support some particular facet of learning.



Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

In the former case, discourse data and learning outcomes are mined for patterns indicative of particular associations between discourse use and learning. In the latter, prior educational research is applied using new analytic techniques. As Cooper (2012) notes, both types are important contributions to the field of analytics. As Gibson (2013) points out (**Table 3**), data driven approaches take a different approach, with potential of such models to simulate sophisticated aspects of higher order learning (with commensurate ethical concerns with auto-classifying students (Johnson, 2013)). In such a view, inductive (“data-driven”) approaches are used to derive data, from which hypotheses can be produced (validated), while deductive processes (the traditional scientific hypothetico-deductive model) approaches construct hypotheses to test through data collection and analysis.

**Table 3 - Comparison between traditional and data-driven science methods (from Gibson, 2013, p. 6)**

	Traditional scientific method	Data-driven science method
Step 1	Ask a Question	Ask a Question
Step 2	Do Background Research	Do Background Research
Step 3	<i>Construct a Hypothesis</i>	<i>Simulate Theory or Search for Patterns</i>
Step 4	<i>Test Hypothesis with Experiments</i>	<i>Generate and Analyze Data</i>
Step 5	<i>Analyze Data and Draw a Conclusion</i>	<i>Conduct Validation Experiments</i>
Step 6	Communicate Results	Communicate Results

In both cases, the relationship between prior theorising and empirical work, and new techniques should be established. Validation is fundamental here, and understanding the ways in which concepts are operationalised across the various manual, semi-manual and automated methods is important (see for example, Chi, 1997). Of course, there are also cases where modelling computationally can inform theory. To give an example, Reimann (2009) discusses process models which analyse sequences of events – of relevance here insofar as they may be identified automatically from log data – in contrast to more traditional approaches which seek to model the variance in the occurrence of dependent variables. In this discussion, Reimann notes that:

Process mining can serve a number of purposes, among them: (a) Discovery—No a priori model exists. Based on an event log, a model is constructed; (b) Conformance—An a priori model exists. Event logs are used to determine the extent to which the enacted collaboration corresponds to the model; (c) Extension—An a priori model exists. The goal is not to test but to extend the model, for instance with performance data (e.g., durations of activities). (Reimann, 2009, p. 251)

As Reimann notes in later work:

For [Educational Data Mining] EDM to contribute more directly to theory building, it needs to be applied to data that measures theoretically relevant properties and mechanisms. These are not

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

necessarily found in the log files of software that has been designed for practical educational rather than research purposes. It is unlikely that we will find many theoretically interesting relations by looking at more of the essentially same kind of data. “Big data” and “more data” are not identical with conceptually “rich data” and “deep data” that capture not only multilayered phenomena but also a rich account of learning contexts. The latter, we believe, could be a productive methodological direction. (Reimann, Markauskaite, & Bannert, 2014, p. 11)

Crucially, the two aspects of analysis that Chi (1997, p. 311) notes in relation to quantitative and verbal analyses hold true here too: First, generation of the right questions is crucial for determining the types of analyses one conducts; and second, understanding the means (the ‘mechanics’) through which one conducts that analysis is a crucial consideration, which should not be thought of as removed from (nor integrated in) the first. Similarly, we should take lessons about the translation of research to practice, and the transfer of practice from site to site into account in our analytics, bearing in mind that “educational data for learning analytics is context specific and variables carry different meanings and can have different implications across educational institutions and area of studies.” (Ifenthaler & Widanapathirana, 2014, p. 1)

## 2. DCLA – UNDERSTANDING WHERE DISCOURSE MEETS ANALYTICS

The preceding discussion of learning analytics lifecycles, and the role of top down and bottom up processes indicates that, although there is clear interest in the application of existing educational research to online contexts, particularly where analysis and learning support may be automated, more research may be needed to understand the application of such research to online contexts. That is, the application of our existing theory to the context of novel forms of data – the types of data capturable, the contexts in which the data is created and captured, and the differences in the medium of online contexts – may need further research.

To give an example, we have a strong interest in ‘exploratory dialogue’ (see for example, Mercer & Littleton, 2007), which in small group contexts has a demonstrated relationship with improved learning outcomes. However, little research has been conducted on exploratory dialogue in online contexts (see for example, Littleton & Whitelock, 2005), with only a few studies of such asynchronous dialogue (see for example, Ferguson, Whitelock, & Littleton, 2010), and none that we are aware of in the context of large multi-party and multi-modal conference chat systems. Yet, it is clear that there are differences between face to face and online communication, and many online contexts provide different types of opportunities for communication. This is an important point regarding a ‘middle space’ between the learning sciences and analytic techniques (Suthers & Verbert, 2013); even a meeting in the middle space can be approached from either side, and such trajectories affect the ways we action our theories. If our operationalization of prior (offline) educational research is driven by the analytic techniques available, then the types of validation we must subsequently engage in, and the ways in which we consider the on and offline dialogue as related or not, should be considered. Similarly, our analytic techniques should not be limited by the ways in which educational dialogue has previously been operationalised – new

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

analytic techniques afford new opportunities to theorise and reconsider what constitutes productive dialogue.

Indeed, if we consider the classic ‘quantification of verbal analysis’ text (Chi, 1997), in that (manual) case the point is made that all stages of analysis involve both top-down and bottom-up processing; coding is driven by top-down theory, but the application of codes is refined through bottom-up protocols (Chi, 1997, pp. 309–310). To give a concrete example in the machine learning case, Rosé and Tovaes note, “more attention should be given to the problem of representing data appropriately” (Rosé & Tovaes, In Press, p. 6), particularly with regard to the issue of segmentation where different levels of granularity in segmentation for the training and classification of various forms of natural language data may have an impact. This concern is a complex one, and in earlier work Arguello and Rosé (2006) propose that with regard to topic segmentation, researchers first need a definition of ‘topic’ which should be:

1. Reproducible by human annotators
2. Not reliant on domain-specific, or task dependent knowledge
3. Grounded in generally accepted principles of discourse structure (in order that shifts in topic are recognisable from surface characteristics of the dialogue)

Thus, although in that case hidden Markov models are used to detect topic shifts, the point is to reproduce a human-recognised topic model (rather than anything more ephemeral). That is, although ‘bottom up’ methods may create productive means of segmentation, the relationship of such methods – and their outputs – to theoretically grounded assumptions should not be forgotten.

In the next two sub-sections, we give some examples of contexts in which discourse data is produced, and associated with learning outcomes. In the section ‘Discourse sites’, we introduce discourse contexts in which data is collected, and the implications of those learning contexts and the data collected for the analytics we conduct. This section is broadly concerned with the ‘top down’ approach to DCLA, derived from existing learning theory and empirical work. In the second section ‘learning analytics’ we discuss a broad set of analytic techniques for discourse data, and the implications of these for our learning-discourse analysis, and feedback-representations. This section is broadly concerned with the ‘bottom up’ approach to DCLA, derived from rapidly advancing analytic techniques.

## 2.1 Discourse Sites

The aim of this (and the next) section is to introduce some examples of work on the ‘sites’ of discourse noted in **Error! Reference source not found.**, around which there has been interest in developing analytics. We use the examples to illustrate the ways in which people develop their reasoning through dialogue, build arguments in CSCL environments, and make rhetorical moves in written texts. If we return to **Error! Reference source not found.**, readers will recall rows with a focus on:

- 1) The kinds of subject knowledge people evidence
- 2) The kinds of argument skills or dispositions they evidence
- 3) The ways in which learners interact with each other and ranges of documents
- 4) The ways in which meaning is built up through shared interaction on documents

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

What we see from this is that, while identifying subject knowledge (content, entities, concepts, etc.), is important, it is not the only crucial element. So too for interaction – whether with citations, connecting ‘nodes’ in CSCL knowledge-building environments, or simply through response to other’s arguments or moves (e.g. responding to a question). Instead, it is combinations of these important features of a text which are crucial to understanding what is going on. The first two rows of the table then are about individual levels of knowledge, the third about collective levels, and the fourth about collaborative co-construction.

We use the following sections to highlight some exemplar (but by no means exhaustive) learning sciences literature around discourse, taking as our focal point theorising around argumentation and reasoning. In the following paragraphs, we present three types of discourse data: unstructured dialogue; free-text; and dialogue within structured CSCL environments. In each case we are interested in the forms of data available and the contexts in which they are obtained. These are of particular interest for the following section, which discusses the ways in which such data has been treated analytically, and raises some potential areas of interest. The point here is to particularly note how theory informs our collection of, and analysis of, data and to foreground the levels at which DCLA might operate.

### 2.1.1 Unstructured dialogue

Perhaps the most obvious form of discourse data is the kind of computer-mediated-communication seen in ‘chat’ programs, and free form dialogue of classrooms (and indeed, everyday conversation). In everyday classroom contexts, teachers model the use, and pay attention to the uptake of, key content and rhetorical terms in student dialogue (rows 1 and 2 of the table). A common complaint in school classrooms at least, is that whole-class discussion focussed on questioning tends to focus too much on these levels of – individual – content-based understanding, rather than whole class commonality (Alexander, 2008). For these levels of discourse use (the latter 2 rows of the table) classroom ‘assessment for learning’ techniques such as ‘mini-whiteboards’ and interactive whiteboards (Hennessy & Warwick, 2010) for all students to display an answer, or more dialogic modes of discussion (Alexander, 2008), may be more important. In our own work we have a strong interest in the ways in which unstructured dialogue is used to co-construct common knowledge (Edwards & Mercer, 1987; Littleton & Mercer, 2013), contributing to the strong consensus among researchers that in a variety of contexts, high quality dialogue is associated with learning (see the collection edited by Littleton and Howe (2010)). A core component of this research – along with that around Accountable Talk (Michaels, O’Connor, Hall, & Resnick, 2002; L. B. Resnick, 2001) – is the analysis of how ideas (including subject knowledge) are deployed in arguments, and taken up by others, to help build new knowledge together. Fundamental to this process is critical but constructive engagement with others’ ideas, participation from all members, respectful challenges and justifications, and the consideration of opinions before agreement is sought – in short, to make the reasoning visible in the dialogue (Mercer & Littleton, 2007).

### 2.1.2 Written texts

Of course, spoken dialogue through short exchanges is not the only site of learning. In particular in formal learning contexts – although by no means exclusively – much learning also occurs through the

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

construction of, and is represented in, longer free-form written texts. Thus, one means through which the kind of co-construction noted in the 4<sup>th</sup> row occurs, is in the interaction with ‘the given’, with written texts of various forms (see, for example, Knight & Littleton, 2015). Indeed, this too should be a core area of interest for DCLA researchers – how learners engage with texts, to construct longer pieces of writing, and how that writing is a site for, and representation of, their learning. Again, subject knowledge is key here, but so too is genre – including the kinds of ‘rhetorical moves’ (or arguments) available, and so on. As Simsek, et al., (2013, p. 3) note, analysis of texts for factual statements provides only a partial picture; in scientific research publications, authors make connections between claims of various sorts including supporting, refuting, and so on and this meta-discourse is fundamental to analysis of such papers.

Again, we see the importance not only of subject knowledge, but argumentation skill (rhetorical moves made), use of other’s ideas (through citation, and background knowledge), and genre-specific knowledge building (each discipline having its own ‘style’ of making sometimes similar claims). Additional Natural Language Processing (NLP) approaches of interest to DCLA can be found in the emerging field of Argument Mining, which sits at the convergence of machine learning, information retrieval and argumentation theory (Palau & Moens, 2009). Again, the four rows of the table can be explicated in terms of this data – at the first two levels (or rows), the use of particular writing genres or scripts, and deployment of key concept terms, with the third and fourth rows seen in deeper engagement with genres, and deeper interactive reasoning around texts.

### 2.1.3 Structured dialogue and CSCL environments

As we note above, the relationship between ‘given’, or sole-author student created texts, and student dialogue, are not firm. Indeed, there have been calls for “abandoning the forced dichotomy between two genres of collaboration tool” (Enyedy & Hoadley, 2006, p. 414), calling for fusion between information (document), and communication interfaces in part to facilitate the kinds of learning identified in the third and fourth rows of **Error! Reference source not found..** The use of CSCL tools to support document-oriented, or less directed reasoning tasks is widespread; aiming to support the kinds of individual learning identified in the first two rows of **Error! Reference source not found..** Much work has been conducted on developing environments which support, and make available for researchers, particular types of dialogue. Importantly though, while de Vries, Lund and Baker (2002) ), to take one example, note important characteristics of productive epistemic dialogue which bear striking resemblance to those described in this paper, they note that such dialogue is not ‘automatically’ produced in structured environments. Indeed, Dillenbourg (2002) notes that some such environments risk ‘over structuring’ and thus restricting the use of important types of dialogue. We note, then, that while design may reduce computational difficulties (for example, by introducing threading to discussions), the points made in this paper with respect to the importance of context are still fundamental to understanding the dynamic features of dialogue through which learning is co-constructed. Computer environments may be seen as complementary to such dialogue, in particular where they formalise, through the user interface conceptual model, some of the systems through which exploratory and accountable dialogue are more likely to occur – the ‘ground rules’ of each. It is thus that systems have been developed specifically to support particular types of formalised argumentation



Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

schema (Clark et al., 2007; De Liddo, Buckingham Shum, Quinto, Bachler, & Cannavacciuolo, 2011; Scheuer et al., 2010; Weinberger, Ertl, Fischer, & Mandl, 2005; Weinberger & Fischer, 2006). Again, we highlight the range of ways in which CSCL argumentation aligns with the table with some scripting aimed at individual learning gains (particularly around argumentation schema), and other methods aimed to aid in intersubjective meaning making.

## 2.2 Analytic Levels

Above, then, we have highlighted some of the data forms, and contexts, of learning discourse. The aim of this section, then, is to discuss the data-types and learning theory noted above, in light of analytic techniques. We note that, even in contexts in which inductive or exploratory approaches to data analysis are used, the data itself is still imbued with theory insofar as the learning contexts from which that data is collected require theorising – for example, genre detection requires a notion of ‘genre’, and the collection of documents displaying such classes. Indeed, in the rather different context of transcribing audio and video recordings, Hammersley (2010) makes similar points: Transcriptions cannot be said to be purely ‘constructed’, they bear an obvious and undisputable relationship to their object of representation which is closely associated with their theoretical ontology, the prior knowledge of the transcriber, and the purposes for which the transcription is being created; similarly though, they are not ‘given’, direct insights into the world, shared understanding is constructed in the process of transcription itself and relies heavily on a (given) set of cultural and theoretical values including the best ways to represent given classes of behaviour (for example, whether to include notes on intonation). In this section, then, we highlight the ways in which theory should penetrate analytics for DCLA, and the complexity arising from the intersection between our data type (dialogue, text, CSCL) and analytic focus (subject knowledge, argumentation skill, interaction, and broader knowledge-building culture).

### 2.2.1 The individual level

If we consider again **Error! Reference source not found.**, recall that the first two rows regard individual level learning gains – in terms of content or subject based learning, and personal psychological development. Analysis of such learning through discourse focuses on the ways in which individuals use terms, whether in individual or collective contexts; what is key is not the type of discourse data, but rather the analytic focus. The most basic kinds of analysis take simple bag of word, or cue phrase approaches – in which classes of dialogue are assigned based on the presence of key words or phrases – to detection of concepts. Such approaches can be enriched with ontologies to match key-phrases from some domain, with those present in the discourse data. More complex variants on this approach might include a broader feature set – such as grammatical markers particularly in the case of argumentation features – to analyse the presence of particular types of discourse. As such, analysis focuses not on the interactional features of discourse, which such techniques are not methodologically equipped to analyse, but on the individual’s use of terms and styles of discourse, and their relationship to individual learning (of concepts, and perhaps rhetorical moves of some kind, and so on).



Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

This kind of approach can be applied across dialogue, written-text, and CSCL environment discourse data, although the former more naturally lends itself to these techniques. Written essays also offer opportunity for a variety of other semantic-related analyses such as latent semantic analysis (LSA), which goes beyond cue-phrase analysis to look at underlying (latent) semantic content in the text. Such analyses can also be conducted on appropriately segmented dialogue data, for example, Epistemic Network Analysis (Shaffer et al., 2009) conducts a principal component analysis to explore connections between themes throughout a dialogue. A major advantage of CSCL environments for such analysis is the ability to build in ‘tagging’ systems to enable learners to ‘tag’ their entries with semantic labels, which can then be analysed. In each of these cases feedback can focus on the types of *concepts* and *facts* included in the discourse, and their incidence. Similarly, analysis for argumentation, rhetorical moves, and tendency (or disposition) to use such devices can also be analysed through these methods. Again, feedback can focus on incidence of particular markers, but it can also – when combined with content analysis – focus on the ways concepts and facts are connected using argument-markers (for example, “it is argued that Broca’s area is related to language because evidence from studies (ref1, ref2, etc.)...”, where Broca’s and the references noted are conceptual or factual claims, and the marker ‘because’ links them). Again, CSCL knowledge-building/argument mapping environments are particularly well equipped for this kind of analysis, in which the ‘connections’ learners make between claims (often represented as ‘nodes’) are often pre-labelled, and thus in machine-readable format.

These notions then cover the ways in which individuals use language to share meaning, and express their knowledge. However, we are also interested in how learners interact with each other using dialogue, how they move beyond *expressing* to ‘taking up’. There are two levels of interest here, the first (and third row of the table) regards the ways in which learners interact with each other and ranges of documents, how they pass information to each other, which is then used in the collective context. The second regards the ways in which meaning is built up through shared interaction, how knowledge is co-constructed. Again, the first two rows of the table then are about individual levels of knowledge, the third about collective levels, and the fourth about collaborative co-construction.

### 2.2.2 The collective small group level

At the collective level of analysis, the focus of attention shifts from expressions by individuals, to the ways in which individuals – and the things that they say – interact. Thus, for example Rosé and colleagues have explored transactivity (sometimes called intertextuality) in small groups (Sionti, Ai, Rosé, & Resnick, 2011; Stahl & Rosé, 2011), dyads (Gweon, Jun, Lee, Finger, & Rosé, 2011), whole-class discussions (Ai, Sionti, Wang, & Rosé, 2010) and its use for summarising group discussions (Joshi & Rosé, 2007). Similarly, the ways in which written texts can be used to resource further written texts - in or out of relatively structured environments - is clearly important (Enyedy & Hoadley, 2006; Knight & Littleton, 2015; Lid & Suthers, 2003; Suthers, Vatrappu, Medina, & Dwyer, 2007). Again, the structured nature of CSCL environments supports analysis to explore the ways in which ideas are ‘taken up’ by others, and knowledge is pooled – collectively – to create (perhaps individual) answers. For example, ‘contingency graphs’ in CSCL environments (Medina & Suthers, 2009), provide a visual insight into the ways in which one ‘idea’ (a node), is ‘contingent upon’ earlier ones. In each of these cases analysis can focus on how

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

terms are stated by an individual or given-text, and then subsequently repeated by other individuals. More sophisticated analysis – particularly in the case of longer written texts – might also focus on the latent-semantic content of a student-produced text in comparison to those from which it is supposed to draw (see for example, Hastings, Hughes, Magliano, Goldman, & Lawless, 2012). Indeed, experimental work can also be modelled on Azmitia and Montgomery (1993), who artificially manipulated which pieces of information participants in a group had, such that information *needed* to be passed between members in order to complete a task.

However, as Stahl (2013) recently noted, in such cases transactivity is being explored at the analytic level of individuals who have (differing) partial information regarding a problem that requires pooled information to solve it, rather than at the group level which has come to be of great interest to the CSCL community. This focus on individuals in collaborative contexts, in contrast to collaborative units, is common to much group-work research, for example in exploring transactivity Azmitia and Montgomery's (1993) interest was in dialogue used to build upon a partner's explicit reasoning statements, rather than on the language used to co-construct knowledge. This, then, brings us to our last interest – that of the co-constructive level (the 4<sup>th</sup> row) of group-cognition.

### 2.2.3 The co-constructive small group level

A fourth level of analysis is at the group-co-constructive level. Indeed, notions of transactivity are highly relevant here, for example, Sionti et al., (2011) summarises notions of transactivity as sharing a requirement for interlocutors reasoning to be explicitly displayed, and usually for contributions between speakers to be connected such that a speakers utterances resource future utterances.

In addition to this interesting construct, it is also worth mentioning other work – relating more closely to active participation and the reasoned consideration of other's arguments – in which the notion of 'heteroglossia' is explored. Heteroglossia (Bakhtin, 1986) is related to the multi-vocality of perspective, the characteristic of a text as displaying, and being open to, multiple views – a significant element of dialogic education (Wegerif, 2011). Building on Martin and White's (2005) description of dialogic expansion (in which alternative positions are available), and dialogic contraction (in which the scope of permitted perspectives is restricted) heteroglossia has recently been operationalized in the computational linguistics context as:

the extent to which a speaker shows openness to the existence of other perspectives apart from the one that is reflected in the propositional content of the assertion being made....Within our Heteroglossia analysis, assertions framed in such a way as to acknowledge that others may or may not agree, are identified as heteroglossic. We describe it as identifying wording choices that do or don't treat other perspectives than what is expressed in the propositional content of the assertion as open for consideration within the continuing discussion. (Rosé & Tovaes, In Press, pp. 10–11)

In the context of a science classroom task, this has been operationalized at the coding level for a four part scheme which takes co-constructive dialogue to be an 'openness to the other' in the way phrases are used. By building in further features at the small-group level, such as looking at dialogue-acts for interaction properties within the group (Erkens & Janssen, 2008; Král & Cerisara, 2012; Stolcke et al., 2000) further conclusions can be drawn with regard to the ways small-groups work together in co-constructive ways, to create new meaning together. In such analytic approaches, the focus is on the

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

ways in which terms are ‘taken up’ by others, and the collective accretion of knowledge over time – as indicated in the fourth row of the table. To analyse such learning, a temporal analysis of features; understanding how terms are used, by whom, and how they relate to each other, is crucial (Mercer, 2008). While temporality is of course of relevance to other areas of learning (including rows 1-3), it is particularly salient to this fourth, co-constructive, level.

## 2.3 Section Summary

In learning contexts we are interested not only in the deployment of particular terms, but of their *effective deployment* in contexts which indicate an understanding of the inferential properties of the term to other concepts. Analytics which explore key words in abstracted ways may obscure the misuse of terms, or – moreover – their simple copying from texts which students use for contextualizing purposes, such as task instructions. The pragmatic level of description is therefore important: syntactic or semantic levels of description can be blind to understanding what is being done in interaction. For example, the ‘same’ question (in syntactic and semantic terms) might be asked at the beginning and end of a lesson, while serving different (pragmatic) functions: in the first instance, to gauge baseline understanding and provide a reference point for the second posing of the question, which is to see how the question may now be reinterpreted. Describing what is happening is not the same as understanding what is being done.

We are mindful that, with regard to the complexity of discourse data, one solution – for example through digital games (Gee, 2008; Shaffer, 2008) – is to ‘design in’ for the types of language we want to teach, and analyse; providing opportunities for their display and their capture in ways that are easy to process by machines. Certainly this point is important, and DCLA should consider the ways it might be informed by, and inform, such structured environments. The potential to build complex, discursive, learning contexts is important and the full breadth of ways in which discourse is constitutive of learning should not be glossed by limiting analytic techniques.

## 3. CONCLUSIONS: WHAT IS DCLA?

There is a growing interest in, and availability of, data and techniques to analyse that data. Discourse is an important feature of learning, and fundamentally associated with the context of social learning analytics, dispositional analytics, and other emerging analytic-sub-fields. We should strive to understand discourse in this wider context. A further driver here is the fact that interest in online text summarisation tools, and support for writing both in and out of formal educational contexts will continue to grow – again, DCLA may play an important role here, perhaps complemented by other forms of analytic tools (such as social network analytics).

### 3.1 A Definition for DCLA

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

This paper has highlighted the ways in which we see DCLA as distinctive, and in doing so, we have tried to contextualise it with respect to the wider research. We began the paper by noting two early focal points for DCLA:

1. Devise and validate analytics that look beyond surface measures in order to quantify linguistic proxies for deeper learning
2. Once researchers have developed and validated discourse-centric analytics, how can these be successfully deployed at scale to support learning?

As we have highlighted throughout this paper, we must be cautious here – linguistic activity is not simply an *indicator* for deeper learning, it is the very site of that learning; when we talk to each other, write texts, engage with knowledge building environments, we are not simply representing what is known, providing *indicators* for that knowledge – we are actively engaging in a co-constructive activity, and that activity constitutes learning. It is with this understanding of high quality learning discourse that we propose a definition of DCLA:

- DCLA focuses on analytics to support high quality discourse for learning contexts; it consists in analysis of discourse data, creation of effective feedback to learners and educators, and the validation and theorising of our analytic techniques.

This definition of course leaves open the particular theorised account of discourse, and pedagogic approach associated with particular types of discourse. Throughout the paper, we have tried to highlight some core pivot points for DCLA, seeking to elaborate our understanding of learning analytics as centered around discourse data. In particular, we have noted the ways in which various sites of discourse might be seen as related to, or constitutive of, learning at a variety of analytic levels – from the individual, to the small group. The definition given does not preclude the use of analytic devices for analysis of learning *indicators* in discourse, of interest particularly to evidence learning at an individual level, but rather it leaves open the space to explore the full range of levels at which learning and discourse are related. It also foregrounds the need for analytics *centered* on discourse to consider validation in transferring analysis from existing learning theory to analytic techniques, and vice-versa, as well as the need to consider the wider learning and feedback systems within which analytic techniques are deployed – particularly as targeted at *learning* outcomes. We close the paper with three key issues which researchers in DCLA will need to further consider: The role of context; DCLA systems; and effective impactful feedback.

### *3.1.1 What is the relationship between discourse in different contexts (particularly on, and offline), and what are the implications of this for detection of particular types of discourse?*

As we note above, the differences between on and offline instantiations of particular discourse-genres are complex, and this area will be of much importance. In order to apply ‘offline’ research in online contexts, we must be sensitive to the differences, and the possible need to ‘validate’ such a shift. Similarly, lessons from online contexts may not translate directly to offline ones – not least because the types of data we can gather are rather different, and thus so too may our insights be. For example creating written texts using technologies including collaboration tools such as etherpads or google docs provides different affordances, and research data, to ‘pen-and-paper’ exercises; or consider the wealth of research on computer-mediated communication (CMC) which recognises, and explores, differences in

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

CMC and ‘offline’ talk. DCLA researchers should consider the role of existing theory, and the place of analytics in that context. This issue thus concerns:

1. The practical-context sensitivity of DCLA and indeed, all discourse analysis with regard to pedagogic, assessment, cultural, and individual differences aspects
2. The theoretical-context sensitivity of DCLA, and how alignments and mismatches between on and offline contexts might be addressed
3. The role of inductive and deductive techniques in advancing analytics and theory
4. The need for validation of learning analytics techniques

### 3.1.2 *What systems need to be in place for the successful deployment of DCLA?*

DCLA rely on particular tools for data capture and analysis. However, their use may also make commitments to particular pedagogic, assessment, and epistemological assumptions (Knight et al., 2014). While tools have particular affordances (for example: twitter replies; forum threading; knowledge mapping connections), these tools can be used in many learning contexts, and their data analysed for many purposes. This highlights that the issue of ‘systems’ not only relates to the analytic device, but also to the importance of meaningful activities in learning contexts for learning analytics. We agree that more attention should be paid to such claims, and that design strategies have strong potential for the development of meaningful, and enjoyable, learning activities centred on language use (Gee, 2008; Shaffer, 2008). Indeed, much work has been conducted on developing environments which support, and make available for researchers, particular types of discourse. DCLA can play a role in this. In considering the necessary systems for the successful deployment of DCLA we should thus consider:

1. The technical factors involved in DCLA, the types of data we can (and want to) collect, and how they are processed
2. The pedagogic and assessment context within which data is collected – collection of data should occur in purposeful, pedagogically valuable contexts
3. An understanding of the contexts within which certain discourse types are likely to occur, such that we do not deploy the ‘wrong’ analytic device in contexts where we would not expect its target class to occur.

### 3.1.3 *What is the most effective way to feedback results to learners, educators, and policy makers for maximum impact?*

As we discuss above, the learning analytics cycle is not solely about the creation of numerical insights into the target data, discourse in DCLA. Rather, we are interested in how we can develop tools to support education. As such, understanding how best to feedback *discourse* supporting data, to whom, in what form, and from what data, is crucial. While computationally complex analytics might hold potential, simpler techniques, might also be impactful. For example, analytics which indicate which topics are being talked about, by whom, giving information such as turn taking — without evaluating ‘good’ or ‘bad’ performance — might provide pedagogic support, while being very computationally simple. In addition, learners, educators and policy makers each have different needs; analytics may provide representational tools to resource discussion around learning impact at all levels. A part of analytics is creating representational tools to understand data we already have; this question relates to



Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

that data, and new data emerging from learning analytic techniques. Whether computationally simple or complex tools are deployed, interpretive flexibility plays a role. This issue thus concerns:

1. The ways in which the same data can be collected, analysed, and fed back to relevant stakeholders at the macro-meso-and micro levels (see Figure 1 below)
2. Research around visualisation, at the various levels, to support “in the moment” and “after the fact” discourse
3. In addition to supporting learning *through* DCLA, research should also be conducted on how to support learning *about* DCLA – how learning situations can make best use of DCLA and in particular how at each level (macro-meso-micro) stakeholders can communicate with other around DCLA sensemaking (Knight, Buckingham Shum, & Littleton, 2013).

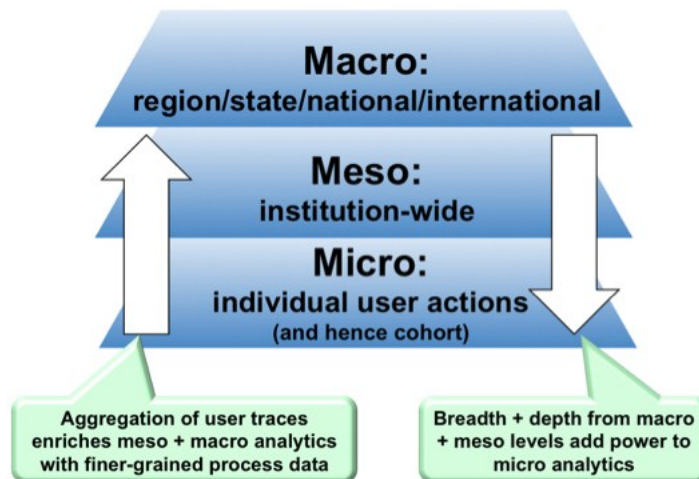


Figure 1 - The macro-meso-micro levels of analysis for learning analytics (from, Buckingham Shum, 2012, p. 3)

We can envisage scenarios, therefore, in which leaders at institutional level and above see aggregated DCLA (for example, as proxies for critical thinking, constructive knowledge building, or successful deployment of online forums), and are able to compare datasets from many contexts. The extent to which comparisons can be made, with integrity, within and between contexts is of course an issue afflicting all types of analytics (educational or otherwise), such is the ease with which digital data may be fused. The particular roles that discourse plays in learning, which we have sought to illuminate in this paper, serve as an alert to the risk of decontextualizing and comparing the different types of DCLA without due care. That being said, there is an intriguing array of opportunities within the middle space of learning sciences and computational techniques around discourse data; by staking the distinctive contribution of DCLA, we highlight this potential, and core challenges for DCLA researchers moving forwards.

#### 4. ACKNOWLEDGEMENTS



Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

We are grateful to Simon Buckingham Shum for feedback on earlier drafts, and to the anonymous reviewers whose comments helped to improve this version. Our thanks to David Gibson and Simon Buckingham Shum for kind permissions to include the table listed as ‘Table 3’ and the figure listed as ‘Figure 1’ respectively. Particular thanks to Tanya Elias for releasing the table listed as ‘Table 2’ in this paper under a Creative Commons BY-NC-ND license.

## 5. REFERENCES

- Ai, H., Sionti, M., Wang, Y.-C., & Rosé, C. P. (2010). Finding transactive contributions in whole group classroom discussions. In *Proceedings of the 9th International Conference of the Learning Sciences - Volume 1* (pp. 976–983). Chicago, Illinois: International Society of the Learning Sciences. Retrieved from <http://dl.acm.org/citation.cfm?id=1854360.1854485>
- Alexander, R. (2008). *Towards Dialogic Teaching: Rethinking Classroom Talk* (4th edition.). Dialogos.
- Arguello, J., & Rosé, C. P. (2006). Topic segmentation of dialogue. In *Proceedings of the HLT-NAACL 2006 Workshop on Analyzing Conversations in Text and Speech* (pp. 42–49). Retrieved from <http://dl.acm.org/citation.cfm?id=1564542>
- Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development*, 2(3), 202–221.
- Bakhtin, M. M. (1986). *Speech genres and other late essays* (Vol. 8). Austin, Texas: University of Texas Press.
- Ball, S. J. (1999). Labour, Learning and the Economy: a “policy sociology” perspective. *Cambridge Journal of Education*, 29(2), 195–206. doi:10.1080/0305764990290203
- Bannert, M., Reimann, P., & Sonnenberg, C. (2013). Process mining techniques for analysing patterns and strategies in students’ self-regulated learning. *Metacognition and Learning*, 1–25. doi:10.1007/s11409-013-9107-6

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

Buckingham Shum, S. (2012). Learning Analytics Policy Brief. UNESCO. Retrieved from

<http://iite.unesco.org/pics/publications/en/files/3214711.pdf>

Buckingham Shum, S., de Laat, M., De Liddo, A., Ferguson, R., Kirschner, P. A., Ravencroft, A., ...

Whitelock, Denise. (2013). 1st International Workshop on Discourse-Centric Learning Analytics.

In *Proceedings of the 3rd International ACM Conference on Learning Analytics and Knowledge*.

Leuven, Belgium: ACM Press: New York. Retrieved from

<http://www.solaresearch.org/events/lak/lak13/dcla13/>

Buckingham Shum, S., de Laat, M., De Liddo, A., Ferguson, R., & Whitelock, D. (2014). 2nd

International Workshop on Discourse-Centric Learning Analytics. In *Proceedings of the 4th*

*International ACM Conference on Learning Analytics and Knowledge*. Indianapolis, US: ACM

Press: New York. Retrieved from <http://dcla14.wordpress.com/>

Buckingham Shum, S., & Ferguson, R. (2012). Social Learning Analytics. *Educational Technology & Society*, 15(3), 3–26.

Chi, M. T. H. (1997). Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. *Journal of the Learning Sciences*, 6(3), 271–315. doi:10.1207/s15327809jls0603\_1

Clark, D. B., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic Frameworks for Assessing Dialogic Argumentation in Online Learning Environments. *Educational Psychology Review*, 19(3), 343–374. doi:10.1007/s10648-007-9050-7

Clow, D. (2012). The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 134–138). Retrieved from <http://dl.acm.org/citation.cfm?id=2330636>

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

Cooper, A. (2012). *A Framework of Characteristics for Analytics* (A Briefing Paper No. Vol 1., No. 7).

JISC CETIS. Retrieved from <http://publications.cetis.ac.uk/wp-content/uploads/2012/12/A-Framework-of-Characteristics-for-Analytics-Vol1-No7.pdf>

De Liddo, A., Buckingham Shum, S., Quinto, I., Bachler, M., & Cannavacciuolo, L. (2011). Discourse-centric learning analytics. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 23–33). New York, NY, USA: ACM.

doi:10.1145/2090116.2090120

De Liddo, A., Sándor, Á., & Buckingham Shum, S. (2012). Contested Collective Intelligence: Rationale, Technologies, and a Human-Machine Annotation Study. *Computer Supported Cooperative Work (CSCW)*, 21(4-5), 417–448. doi:10.1007/s10606-011-9155-x

De Vries, E., Lund, K., & Baker, M. (2002). Computer-Mediated Epistemic Dialogue: Explanation and Argumentation as Vehicles for Understanding Scientific Notions. *Journal of the Learning Sciences*, 11(1), 63–103. doi:10.1207/S15327809JLS1101\_3

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. *Three Worlds of CSCL. Can We Support CSCL?*, 61–91.

Edwards, D., & Mercer, N. (1987). *Common knowledge: the development of understanding in the classroom*. London, UK: Routledge.

Elias, T. (2011). Learning analytics: Definitions, processes and potential. Retrieved from <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>

Enyedy, N., & Hoadley, C. M. (2006). From dialogue to monologue and back: Middle spaces in computer-mediated learning. *International Journal of Computer-Supported Collaborative Learning*, 1(4), 413–439. doi:10.1007/s11412-006-9000-2

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols.

*International Journal of Computer-Supported Collaborative Learning*, 3, 447–470.

doi:10.1007/s11412-008-9052-6

Ferguson, R., Whitelock, D., & Littleton, K. (2010). Improvable objects and attached dialogue: new literacy practices employed by learners to build knowledge together in asynchronous settings.

*Digital Culture & Education*, 2(1), 103–123.

Furberg, A., & Ludvigsen, S. (2008). Students' Meaning-making of Socio-scientific Issues in Computer Mediated Settings: Exploring learning through interaction trajectories. *International Journal of Science Education*, 30(13), 1775–1799. doi:10.1080/09500690701543617

Gee, J. P. (2008). Learning and games. In K. Salen (Ed.), *The ecology of games: Connecting youth, games, and learning* (Vol. 3, pp. 21–40). Cambridge MA: MIT Press. Retrieved from <http://mitpress2.mit.edu/books/chapters/0262294249chap2.pdf>

Gibson, D. (2013). Psychometric Considerations for Digital Assessments. In *X World Conference on Computers in Education*. Torun, Poland. Retrieved from [http://edu.rsei.umk.pl/wcce2013/publications/v1/V1.27\\_008-Gibson-fullR-FPR.pdf](http://edu.rsei.umk.pl/wcce2013/publications/v1/V1.27_008-Gibson-fullR-FPR.pdf)

Gweon, G., Jun, S., Lee, J., Finger, S., & Rosé, C. P. (2011). A framework for assessment of student project groups on-line and off-line. In *Analyzing Interactions in CSCL* (pp. 293–317). Springer. Retrieved from [http://link.springer.com/chapter/10.1007/978-1-4419-7710-6\\_14](http://link.springer.com/chapter/10.1007/978-1-4419-7710-6_14)

Hammersley, M. (2010). Reproducing or constructing? Some questions about transcription in social research. *Qualitative Research*, 10(5), 553–569. doi:10.1177/1468794110375230

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

Hastings, P., Hughes, S., Magliano, J. P., Goldman, S. R., & Lawless, K. (2012). Assessing the use of multiple sources in student essays. *Behavior Research Methods*, 44(3), 622–633.

doi:10.3758/s13428-012-0214-0

Haythornthwaite, C., & de Laat, M. (2010). Social Networks and Learning Networks: Using social network perspectives to understand social learning. In *7th International Conference on Networked Learning*. Retrieved from [http://celstec.org.uk/system/files/file/conference\\_proceedings/NLC2010\\_Proceedings/abstracts/PDFs/Haythornwaite.pdf](http://celstec.org.uk/system/files/file/conference_proceedings/NLC2010_Proceedings/abstracts/PDFs/Haythornwaite.pdf)

Hennessy, S., & Warwick, P. (2010). Research into teaching with whole-class interactive technologies. *Technology, Pedagogy and Education*, 19(2), 127–131. doi:10.1080/1475939X.2010.491211

Ifenthaler, D., & Widanapathirana, C. (2014). Development and Validation of a Learning Analytics Framework: Two Case Studies Using Support Vector Machines. *Technology, Knowledge and Learning*, 1–20. doi:10.1007/s10758-014-9226-4

Johnson, J. A. (2013). *Ethics of Data Mining and Predictive Analytics in Higher Education* (SSRN Scholarly Paper No. ID 2156058). Rochester, NY: Social Science Research Network. Retrieved from <http://papers.ssrn.com/abstract=2156058>

Joshi, M., & Rosé, C. P. (2007). Using Transactivity in Conversation for Summarization of Educational Dialogue. *SLaTE Workshop on Speech and Language Technology in Education*. Retrieved from [http://www.isca-speech.org/archive\\_open/archive\\_papers/slate\\_2007/sle7\\_053.pdf](http://www.isca-speech.org/archive_open/archive_papers/slate_2007/sle7_053.pdf)

Knight, S. (2013). Creating a supportive environment for classroom dialogue. In S. Hennessy, P. Warwick, L. Brown, D. Rawlins, & C. Neale (Eds.), *Developing interactive teaching and learning using the IWB*. Open University Press.

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

Knight, S., Buckingham Shum, S., & Littleton, K. (2013). Collaborative Sensemaking in Learning Analytics. In *CSCW and Education Workshop*. San Antonio, Texas, USA.

Knight, S., Buckingham Shum, S., & Littleton, K. (2014). Epistemology, assessment, pedagogy: where learning meets analytics in the middle space. *Journal of Learning Analytics*, In–press.

Knight, S., & Littleton, K. (2013). Discourse, computation and context – sociocultural DCLA revisited. In *1st International Workshop on Discourse-Centric Learning Analytics*. Leuven, Belgium.  
Retrieved from <http://oro.open.ac.uk/36640/>

Knight, S., & Littleton, K. (2015). Thinking, Interthinking, and Technological Tools. In R. Wegerif, J. C. Kaufman, & L. Li (Eds.), *The Routledge International Handbook of Research on Teaching Thinking*. Routledge.

Král, P., & Cerisara, C. (2012). Dialogue act recognition approaches. *Computing and Informatics*, 29(2), 227–250.

Lid, V., & Suthers, D. D. (2003). Supporting online learning with an artifact-centered cross-threaded discussion tool. In *Proceedings of the International Conference on Computers in Education* (pp. 319–314). Retrieved from  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.70.8914&rep=rep1&type=pdf>

Littleton, K., & Howe, C. (2010). *Educational dialogues: understanding and promoting productive interaction*. Abingdon, Oxon: Routledge.

Littleton, K., & Mercer, N. (2013). *Interthinking: putting talk to work*. London: Routledge.

Littleton, K., & Whitelock, D. (2005). The negotiation and co-construction of meaning and understanding within a postgraduate online learning community. *Learning, Media and Technology*, 30(2), 147–164.



Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

Martin, J. R., & White, P. R. (2005). *The language of evaluation*. Palgrave Macmillan Basingstoke and New York. Retrieved from

<http://students.lti.cs.cmu.edu/11899/files/LanguageofEvaluationBook.pdf>

Medina, R., & Suthers, D. D. (2009). Using a contingency graph to discover representational practices in an online collaborative environment. *Research and Practice in Technology Enhanced Learning*, 4(03), 281–305.

Mercer, N. (2008). The seeds of time: why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences*, 17(1), 33–59. doi:10.1080/10508400701793182

Mercer, N., & Littleton, K. (2007). *Dialogue and the Development of Children's Thinking: A Sociocultural Approach* (New edition.). Routledge.

Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. (2002). Accountable talk: classroom conversation that works. *Pittsburg: University of Pittsburgh*.

Mislevy, R., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: evidence-centred design, psychometrics, and educational data mining. *Journal of Educational Data Mining*, 4(1), 11–48.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, 60(5), 903–921. doi:10.1007/s11423-012-9265-2

Palau, R. M., & Moens, M.-F. (2009). Argumentation mining: the detection, classification and structure of arguments in text. In *Proceedings of the 12th international conference on artificial intelligence and law* (pp. 98–107). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1568246>

Plato. (1999). *Phaedrus*. (B. Jowett, Trans.). Retrieved from <http://www.gutenberg.org/ebooks/1636>

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CACL research. *International Journal of Computer-Supported Collaborative Learning*, 4(3), 239–257. doi:10.1007/s11412-009-9070-z

Reimann, P., Markauskaite, L., & Bannert, M. (2014). e-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, n/a–n/a. doi:10.1111/bjet.12146

Resnick, L. B. (2001). Making America smarter: The real goal of school reform. *Developing Minds: A Resource Book for Teaching Thinking*, 3–6.

Rosé, C. P., & Tovaes, A. (In Press). What Sociolinguistics and Machine Learning Have to Say to One Another about Interaction Analysis. In L. Resnick, C. Asterhan, & S. Clarke (Eds.), *Socializing Intelligence Through Academic Talk and Dialogue*. Washington, DC: American Educational Research Association.

Rosé, C. P., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 3(3), 237–271.

Rosen, D., Miagkikh, V., & Suthers, D. D. (2011). Social and semantic network analysis of chat logs. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 134–139). New York, NY, USA: ACM. doi:10.1145/2090116.2090137

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, 5(1), 43–102.

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

Shaffer, D. W. (2008). *How Computer Games Help Children Learn*. Palgrave Macmillan.

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., ... Mislevy, R. (2009).

Epistemic network analysis: A prototype for 21st-century assessment of learning. Retrieved from <http://www.mitpressjournals.org/doi/abs/10.1162/ijlm.2009.0013>

Sie, R. L. L., Ullmann, T. D., Rajagopal, K., Cela, K., Bitter–Rijkema, M., & Sloep, P. B. (2012). Social network analysis for technology–enhanced learning: review and future directions. *International Journal of Technology Enhanced Learning*, 4(3), 172–190. doi:10.1504/IJTEL.2012.051582

Simsek, D., Buckingham Shum, S., Sandor, A., De Liddo, A., & Ferguson, R. (2013). XIP Dashboard: visual analytics from automated rhetor-ical parsing of scientific metadiscourse. In *1st International Workshop on Discourse-Centric Learning Analytics*. Leu. Retrieved from <http://libeprints.open.ac.uk/37391/1/LAK13-DCLA-Simsek.pdf>

Sionti, M., Ai, H., Rosé, C. P., & Resnick, L. (2011). A framework for analyzing development of argumentation through classroom discussions. *Educational Technologies for Teaching Argumentation Skills, Bentham Science*. Retrieved from <http://www.cs.cmu.edu/~huaai/pub/Sionti.pdf>

Stahl, G. (2010). GROUP COGNITION AS A FOUNDATION FOR THE NEW SCIENCE OF LEARNING. In M. S. Khine & I. M. Saleh (Eds.), *New Science of Learning: Cognition, Computers and Collaboration in Education*. London: Springer.

Stahl, G. (2013). Transactive discourse in CSCL. *International Journal of Computer-Supported Collaborative Learning*, 8(2), 145–147. doi:10.1007/s11412-013-9171-6

Stahl, G., & Rosé, C. P. (2011). Group Cognition in Online Teams. *Theories of Team Cognition: Cross-Disciplinary Perspectives*. New York, NY: Routledge/Taylor & Francis. Web: <http://GerryStahl>.

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

- Net/pub/gcot. Pdf*. Retrieved from  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.190.3215&rep=rep1&type=pdf>
- Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., ... Meteor, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3), 339–373.
- Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. In *System Science (HICSS), 2012 45th Hawaii International Conference on* (pp. 3368–3377). Retrieved from [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6149232](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6149232)
- Suthers, D. D., Lund, K., Rosé, C. P., Teplov, C., & Law, N. (Eds.). (2013). *Productive Multivocality in the Analysis of Group Interactions* (Vol. 15). Springer. Retrieved from  
<http://www.springer.com/education+%26+language/learning+%26+instruction/book/978-1-4614-8959-7>
- Suthers, D. D., & Medina, R. (2010). The Temporal Development of Representational Practices: Implications for Theory and Analysis of Situated Learning. In *2010 43rd Hawaii International Conference on System Sciences (HICSS)* (pp. 1–10). doi:10.1109/HICSS.2010.397
- Suthers, D. D., Vatrappu, R., Medina, R., & Dwyer, N. (2007). Information sharing and interaction in collaborative convergence. *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS*, 162, 167.
- Suthers, D. D., & Verbert, K. (2013). Learning analytics as a “middle space.” In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 1–4). New York, NY, USA: ACM. doi:10.1145/2460296.2460298

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

Tablan, V., Roberts, I., Cunningham, H., & Bontcheva, K. (2013). GATECloud.net: a platform for large-scale, open-source text processing on the cloud. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 371(1983), 20120071.  
doi:10.1098/rsta.2012.0071

Wegerif, R. (2011). Towards a dialogic theory of how children learn to think. *Thinking Skills and Creativity*, 6(3), 179–190. doi:10.1016/j.tsc.2011.08.002

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer supported collaborative learning. *Instructional Science*, 33(1), 1–30. doi:10.1007/s11251-004-2322-4

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46(1), 71–95.  
doi:10.1016/j.compedu.2005.04.003

Wells, G. (1999). *Dialogic inquiry: towards a sociocultural practice and theory of education*. Cambridge University Press.

Wells, G., & Claxton, G. (2002). *Learning for life in the 21st century: sociocultural perspectives on the future of education*. Wiley-Blackwell.